Full Length Research Paper

Stream Flow Simulation using SVM, ANFIS and NAM Models (A Case Study)

Elham Kakaei Lafdani1, Alireza Moghaddam Nia2*, Azadeh Ahmadi3, Milad Jajarmizadeh4, Mohsen Ghafari Gosheh5

1MSc Graduate of Watershed Management, University of Zabol, Zabol, Iran
2 Associate Professor of Mountain and Arid Regions Reclamation Department, Faculty of Natural Resources, University of Tehran, Karaj, Iran
3 Assistant Professor of Civil Engineering Department, Isfahan University of Technology, Isfahan, Iran
4 Department of Hydraulic and Hydrology, Universiti Teknologi Malaysia, Johor, Malaysia
5Natural Resources and Watershed Organization of Charmahal Bakhtiyari Province, Charmahal Bakhtiyari, Iran
*Corresponding Author: a.moghaddamnia@ut.ac.ir

Stream flow forecasting can be an appropriate indicator in estimating future conditions for water resources management. The present study aimed to compare the efficiency of Support Vector Machine (SVM), Adaptive Neural Fuzzy Inference Systems (ANFIS) and conceptual hydrological model of MIKE11/NAM in simulating the daily stream flow. The studied area is Eskandari basin located in Iran. For this purpose, a ten-year period (1999-2009) of daily data including rainfall, runoff, temperature and evaporation were used. Furthermore, the performances of the models in flow simulation were investigated using statistical indicators of correlation coefficient ($R^2$), Root Mean Square Error (RMSE) and the Nash-Sutcliffe (NS) coefficient. The results showed that every three models possess an appropriate performance and efficiency in the studied area. During testing (verification) period, SVM with the highest correlation coefficient ($R^2=0.99$) and lowest RMSE equal to (RMSE=2.13 $m^3/s$), had a better performance than ANFIS model ($R^2=0.82$, RMSE=3.21 $m^3/s$ ) and NAM model ($R^2=0.75$, RMSE=3.48 $m^3/s$ ). In addition, Nash-Sutcliffe coefficient for SVM, ANFIS and NAM models were 0.99, 0.79 and 0.70, respectively.

Key words: Stream Flow Simulation; SVM; ANFIS; MIKE11/NAM Model.

1. INTRODUCTION

Stream Flow Simulation; SVM; ANFIS; MIKE11/NAM Model.

Water resources management is directly related to the spatial examination of hydrologic cycle (Jajarmizadeh et al., 2012). On the other hands, one of the necessities in watershed management is to utilize the stream flow simulating models in developing operation policies. Many various models are used for modeling this process. These models can be divided into conceptual, physical and data driven models.in recent decades, artificial intelligence techniques such as Adaptive Neural Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN) and Support Vector Machine (SVM) have been widely considered owing to their flexibility in modeling the non-linear processes such as rainfall-runoff model (Elom et al., 2012). Nedbors-Afstronings Model (NAM), the equivalent of Rainfall-Runoff Model in Danish, is an integrated and conceptual model of rainfall-runoff which simulates surface flow, subsurface and base flow (DHI, 1999). NAM model has been applied as a conceptual model of rainfall-runoff in different parts of the world including Bangladesh, Malaysia, Vietnam and etc (Butts et al., 2004; Madsen, 2000; Lippiwattanakarn et al., 2004a). Shamsdin and Hashim (2002) applied NAM model for predicting the flow discharge in Liang River located in northern part of Malaysia. Results show predicted amounts by the NAM model were in accordance with the historical data appropriately. Lippiwattanakarn et al. (2004b) compared back propagation neural network simulation model (BPNN) and NAM models for simulating the flow discharge in Thailand. The results showed that the both models were able to simulate flow reasonably. Liu and Sun (2010) suggested a new modern sensitive analysis scheme for NAM model that indicated the sensitivity analysis problem in a general multi-objective framework.

So many studies have been conducted using ANFIS (Halid nd Ridd, 2002; Navak et al., 2004; Aldrian and Djamil, 2008; Dastorani et al., 2010) and SVM models (Han et al., 2002; Moghaddamnia et al., 2009) all around the world. Aqil et al. (2007) conducted a comparative study
between ANFIS and ANN in modeling continuous daily and hourly behavior of runoff in both daily and hourly time scales. Their results showed that ANFIS had a higher efficiency than other two techniques for forecasting the runoff based on rainfall. Ramesan et al. (2009) investigated the rainfall-runoff prediction in Brue Basin in South West England using NNARX (Neural Network Auto Regressive with Exogenous Input), ANFIS, LLR (Local Linear Regression) and NW (Neuro-Wavelet model). The results indicated that in identical condition in terms of combination of input during the verification model, NW had a better efficiency than other models in the runoff simulation. Han et al. (2007) used SVM for predicting flood in a basin in which the developed model were compared with basic models such as simple model, trend, multiple regression and multivariate neural networks. Kakaei Lafdani et al. (2013) in a study reviewed the functioning of the SVM and ANN models in predicting the amount of daily suspended sediment. Their results indicated high efficiency of these two models in predicting the amount of suspended sediment volume.

In this study, which has been done on Eskandari Basin in Iran, SVM, ANFIS and NAM were examined in daily stream flow simulation.

2. MATERIALS AND METHODS

2.1. Support Vector Machine (SVM)

Support vector machines are supervised learning machines i.e. they are applied for categorization and regression issues. This model has been proposed by Vapnik et al. (1995) and developed according to the Computational Learning Theory (Vapnik, 1998). The structure of SVM model is shown in Figure 1. Similarly, the data are available in a linear equation:

$$y = \text{sign} \left( \sum_{i=1}^{N} y_i \alpha_i (X \cdot X_i) + b \right)$$  \hspace{1cm} (1)

Where $Y$ is the equation output, $y_i$ is the value of training category and $X_i$ is the inner product. $x=(x_1, x_2, ..., x_n)$ Indicates an input data and vectors $X_i$ and $i=1,...,N$ are the support vectors. In equation (8), $\alpha_i, b$ parameters determine the hyperplane. If the data are not separable linearly, equation (1) is changed to the following equation:

$$y = \text{sign} \left( \sum_{i=1}^{N} y_i \alpha_i K(X, X_i) + b \right)$$  \hspace{1cm} (2)

Function $K(X, X_i)$, is a kernel function which establishes inner product for creating machines with different non-linear decision-making levels in data space. In a regression model of SVM, functional dependence of the dependent variable of $y$ is required to be estimated to the set of dependent variables of $x$. two samples of SVM are defined as follows: A) regression models of SVM which are known as $\nu-SVM$, B) Regression models of SVM which are known as $\varepsilon-SVM$. Four standard conversions of kernel function which are often used in modeling and regression are linear, polynomial, Radial Basis Function (RBF) and exponential RBF. In various studies, RBF kernel has been reported as the best kernel. The RBF kernel is shown in equation (3).
In this study, ε-SVM has been used with RBF kernel function. For more information regarding SVMs and types of kernel functions, please see the conducted studies (Kakaei Lafdani et al., 2013; Sivapragasam et al., 2001; Dibike et al., 2001; Han and Yang, 2001).

2.2. Artificial Neural Fuzzy Inference System (ANFIS)

In this part, a brief description of ANFIS structure principles is presented. The structure of ANFIS model is shown in Figure 2. Generally, the ANFIS structure is composed of five layers.

Layer 1 (input conjunctions): Every node in this layer is a square node with following function:

\[ Q_i^1 = \mu_A(x) \]  

Where \( x \) is the input node and \( A_i \) is an expression statement (small, large and etc.) which contributes in this node function. On the other hand, \( Q_i^1 \) is the membership function and this is the degree where sufficiency of \( x \) is determined for \( A_i \). Usually, \( \mu_A(x) \) is determined as bell-shaped with maximum and minimum zero using the following equation:

\[ \mu_A(x) = \frac{1}{1 + \left[ \frac{x-c_i}{a_i} \right]^2} \]  

Where \( \{a_i, b_i, c_i\} \) are the parametric set. Similar to the values of these parameters, its bell-shaped functions vary accordingly; hence, various membership functions of the linguistic indicators are illustrated in \( A_i \).

Layer 2: Every node (conjunction) in this layer is a circle node marked with \( II \) which multiples the input signals and removes the obtained multiplication. Every output node shows the weight rule. For example:

\[ w_i = \mu_A(x) \times \mu_B(y), \quad i = 1,2 \]  

Layer 3: Every node in this layer is a circle node marked with \( N \). \( i \) th node is calculated relative to the fire power of \( i \) th rule from the sum of all weight of the rules. According to this equation, the output of this layer is called normalized weighted rules.

\[ \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,2 \]  

Layer 4: Every node in this layer is a square with following equation:

\[ Q_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \]  

Where \( \bar{w}_i \) is the output of the third layer and \( \{p_i, q_i, r_i\} \) are the parameters set.

Layer 5: The single node in this layer is a circle node marked with \( \Sigma \) which calculates the total output as the sum of all input signals and so forth.

\[ Q_i^5 = \text{overall output} = \sum_i \bar{w}_i f_i = \sum_i \frac{w_i f_i}{\sum w_i} \]  

2.3. MIKE11/NAM model

Fig. 2: The structure of ANFIS model (Remesan et al., 2009)
The hydrological model of NAM is an integrated and conceptual model of rainfall-runoff. NAM model simulates the rainfall-runoff process using the linkage rule between the four reservoirs which are connected together and each is the representative of different physical specifications. These four reservoirs are: snow storage, surface storage, groundwater storage, root zone storage. The required basic data for NAM model are: model parameters, initial conditions, meteorological data and data for hydrometric calibration and validation of the model. Basic meteorological data include precipitation, potential evapotranspiration, wherein snowmelt also is modeled, temperature and radiation data should also be added. In addition, NAM model has the ability of simulating the changes made by human in hydrologic cycle (e.g. irrigation and wells pumping), meanwhile time series of irrigation and using rate of groundwater aquifers will be required. Model parameters include: the maximum of water amount in the reservoir surface area ($U_{\text{max}}$), the maximum amount of water in the root zone storage ($L_{\text{max}}$), overland flow runoff coefficient (CQOF), the time constant for intermediate flow (CKIF), the time constant for the flow routing of interflow and overland (CK12), root zone threshold value for the time constant for base flow (TOF) and time constant for routing base flow (CKBF). The structure of NAM model is shown in Figure 3.

This basin with an area of 1836.95 $Km^2$ is located in the north part of the Zayandehrood dam basin in Iran and its most important river is the Pelasjan. This river plays a leading role in the draining of Eskandari station and almost all the tributaries join this river before reaching the dam's lake. This basin is located northwest of the Zayandehrood dam between the longitudes of 50°2' to 50°41' and the latitudes of 32°12' to 32°46' (Figure 4). Most of the surface of this basin consists of flat lands with a climate varying from very humid in the upstream parts of the basin to semi-humid in the exit part but the dominant climate is semi-humid.

3. RESULTS AND DISCUSSION

In this study, statistical periods of 1999-2006 and 2006-2009 were selected as the training (calibration) period and testing (verification) period of, respectively. The input models was combination of rainfall ($R_t$), temperature ($T_t$), evaporation ($EVAP_t$), and flow discharge on the day before ($Q_{t-1}$) were considered as ($R_t,T_t,EVAP_t,Q_{t-1}$). For the flow simulation using ANFIS model, different membership functions were studied and the results showed that triangular membership function had the best performance compared with other membership functions.
In order to simulation using $\varepsilon-SVM$ and RBF kernel, the optimal values of $C$, $\varepsilon$, and $\gamma$ must be determined. Parameters $C$ and $\varepsilon$ influence on quality and time of training. In addition, parameter $\gamma$ influences on overfitting and underfitting of the network. In this study, the values of these parameters were determined using trial and error method. In selecting the optimal values of the parameters, it was tried that model had the minimal error during the testing phase. Parameters $C$, $\varepsilon$ and $\gamma$ were selected equal to 9, 0.009 and 1, respectively.

Moreover, for the stream flow simulation using NAM model, evaporation, temperature and daily rainfall data into the model were used and the daily output discharge from the basin were also used (to compare the observed and simulated flow). The used stations in Eskandari basin are five stations three of which are rain gauge stations (Eskandari, Fereydan and Bouin), one is Singerd climatological station and another one is hydrometric station (Eskandari). The homogeneity of the rain gauge stations was assessed using double mass curve test. Furthermore, the impact factor of each rain gauge station in the runoff simulation was also calculated according to the basin weighted average rainfall obtained through Thiessen polygon technique.

In this study Root Mean Square Error (RMSE), correlation coefficient ($R^2$) and Nash-Sutcliffe (NS) statistical criteria have been used to review the performance of the studied models. These coefficients are calculated according to the following equations:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Q_{\text{obs}}(i) - Q_{\text{sim}}(i))^2}$$  \hspace{1cm} (10)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Q_{\text{obs}}(i) - Q_{\text{sim}}(i))^2}{\sum_{i=1}^{n} (Q_{\text{obs}}(i) - \bar{Q}_{\text{obs}})^2}$$ \hspace{1cm} (11)

$$NS = 1 - \frac{\sum_{i=1}^{n} (Q_{\text{sim}}(i) - Q_{\text{obs}}(i))^2}{\sum_{i=1}^{n} (Q_{\text{obs}}(i) - \bar{Q}_{\text{obs}})^2}$$ \hspace{1cm} (12)

In above equations, $Q_{\text{obs}}(i)$ is the observed value at time $i$, $Q_{\text{sim}}(i)$ is the simulated value at time $i$, $n$ is the sum number of observations, and $Q_{\text{obsave}}$ and $Q_{\text{simave}}$ are the average of observed and predicted values, respectively. The obtained results from runoff simulation using SVM, ANFIS and NAM models are shown in Table 1.

Given the obtained results in Table 1, it is indicated that every three models have an appropriate performance in runoff simulation in the area. In training (calibration) period, ANFIS models, with highest correlation coefficient (0.98) and lowest RMSE equal to (0.74 $m^3/s$) and highest Nash-Sutcliffe coefficient ($NS = 0.98$), had a better performance than SVM model with ($R^2 = 0.97$, $RMSE = 1.11 (m^3/s)$, $NS = 0.97$) and NAM model ($R^2 = 0.78$, $RMSE = 2.78 (m^3/s)$, $NS = 0.70$).

However, in the verification period, SVM model, with highest correlation coefficient and NS (0.99) and lowest RMSE ($2.13 m^3/s$), was the best.
model in flow simulation. In the testing (verification) period, ANFIS also had a higher capability in flow simulation with correlation coefficient equal to 0.82 and RMSE equal to 3.21 (m³/s) compared with NAM model (R² = 0.75, RMSE = 3.48 (m³/s), NS = 0.70). The scatter plot of observed and simulated flow by SVM, ANFIS and NAM models during testing (verification) period is shown in Figure 5.

The total flow volume during the testing (verification) period was equal to 5016.683 (m³/s). SVM, ANFIS and NAM models could simulate the flow discharge equal to 2031.736 (m³/s), 5652.053 (m³/s) and 5438.44 (m³/s), respectively. In other words, SVM model has provided a more accurate estimation of runoff volume during the verification period. Figure 6 indicates the simulated runoff using SVM model during the testing period.

**Table 1: Results of flow simulation using SVM, ANFIS and NAM Models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Training (Calibration)</th>
<th>Testing (Verification)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>RMSE</td>
</tr>
<tr>
<td>NAM</td>
<td>0.78</td>
<td>2.78</td>
</tr>
<tr>
<td>SVM</td>
<td>0.97</td>
<td>1.11</td>
</tr>
<tr>
<td>ANFIS</td>
<td>0.98</td>
<td>0.74</td>
</tr>
</tbody>
</table>

**Fig. 5:** Scatter plot of observed and simulated flow by SVM, ANFIS and NAM models during testing (verification) period
4. CONCLUSION

In this study, hydrological model of NAM, ANFIS and SVM models were used to simulate stream flow discharge in Eskandari basin for a 10-year period. The models input were the combination of rainfall, temperature, evaporation and flow discharge of the day before. The results of the study showed that every three mentioned models had a high ability for runoff simulation. Meanwhile, SVM model in the testing period had the best performance than ANFIS and NAM models. Although the results of this study suggested that SVM and ANFIS had a better performance than the conceptual NAM model, the efficiency of these models (SVM and ANFIS) is depend upon their training. While in conceptual models, if the calibration model is done well, and its parameters are estimated appropriately based on their physical concept and the region’s condition, they can have a better ability in simulation and a more accurate and considerable analysis can be provided on the results of the simulation by the conceptual model.

REFERENCES


